

Reality Check:

Understanding the Limitations and Appropriate Uses of Census- based Data

Affordable Housing Underwriting Forum & Annual Meeting
National Council of Affordable Housing Market Analysts (NCAHMA)
Chicago, IL September 2011

Julia LaVigne, Ribbon Demographics, LLC

www.ribbondata.com

Rick Stein, Urban Decision Group, LLC

www.UrbanDecisionGroup.com

Acknowledgements

- **Ken Hodges, head demographer (Nielsen)**
- **Lynn Wombold, head demographer (ESRI)**

Sources of information:

- **U.S. Census Bureau**
- **Nielsen, Inc.**
- **ESRI**

Why is it important to understand the impact of the 2010 Census and American Community Survey (ACS) ?

Demographic data products provided by *private firms* are built largely on a statistical and geographic foundation provided by the *U.S. Census Bureau*.

Census 2010 is VERY DIFFERENT from Census 2000

- No more “long form”
- Limited amount of information
- Continuous data collection using the American Community Survey (ACS) replaces long form

Census 2010 only has 7 questions

- Name
- Sex
- Age/birth date
- Hispanic ethnicity
- Race
- Relationship to the person filling out the form.
- Other residence, for example, military or college residence, if they sometimes live at another address.

Census 2010 used to produce Summary File 1 (SF1)

- 331 data variables
- More information than 2000 SF1
- Population, Age, Sex, Race, Household, Family, Group Quarters, and Housing Units (including tenure)
- Does **not** include income data
- Very little detail
- <http://factfinder2.census.gov/main.html>
- Quick Table “QT-H2” shows household tenure by age and household size

What is the ACS?

- Ongoing survey sent to 250,000 households on a monthly basis (3 million annually)
- Data used to distribute over \$400 billion in state and federal funds each year
- Some information is the same as 2010 Census data, this includes:
 - Age
 - Sex
 - Race
 - Family & Relationships

What is the ACS?

- Most information is not collected by Census 2010, this includes:
 - Income and benefits
 - Housing Unit Characteristics
 - Where you work and how you get there
 - Where you live and how much you pay for some essentials
 - Disabilities
 - Health insurance
 - Education
 - Veteran status

What is the ACS?

ACS provides up to 3 types of datasets (a total of 9 possible for any given year) to consider when analyzing a particular area.

- ***Single year*** estimates for areas 65,000+ in population
- ***3 year*** estimates for areas 20,000+ in population
- ***5 year*** average of estimates for all Census geographies

ACS Data Collection & Methodology

- Residency rules are different
- Date-specific variables represent monthly averages
- Income adjusted by the CPI to represent a calendar year
- Migration measured from 1 year ago
- Survey samples are weighted by estimates, not counts
- Smaller sample sizes

Which of the 3 datasets should you use?

- **5 year** estimates are more accurate (smaller MOE) but not as current as **1 or 3 year**
- **5 year** estimates are *expected* to be the dominant source of information
- **5 year** estimates are the official replacement for the long form
- **2005-2009** dataset is the first 5 year dataset available
- **2005-2009** dataset is weighted to the 2000 Census
- **2006-2010** will be weighted to the 2010 Census

Concerns with the ACS

- Sample size is smaller than before
- Margin of Error (MOE) is larger than before
 - MOE is now published with the data
- Also subject to significant non-sampling errors that cannot be measured
- 3 Possible contradicting datasets
- Misinterpretation & misuse

Explanation of Terminology

- **Margin of Error** – a measure of the variability of the estimate due to sampling error.
 - 90% chance the number is within the upper and lower bounds of the MOE.
- **Confidence Interval**
 - Upper Bound – estimate *plus* the MOE
 - Lower Bound – estimate *minus* the MOE
- **Coefficient of Variation** – measures the amount of sampling error relative to the size of the estimate

ACS Example (different time periods) – Cook County, IL

Total Households

2009 ACS

	Value	MOE	CV
1 Year	1,931,361	9,291	0.29%
3 Year	1,939,904	5,171	0.16%
5 Year	1,941,417	5,109	0.16%
2010 Census	1,966,356		

ACS Example (different time periods) – Albany, GA

Total Households

2009 ACS

	Value	MOE	CV
1 Year	35,949	1,821	3.08%
3 Year	36,262	1,183	1.87%
5 Year	36,489	853	1.42%
2010 Census	29,781		

ACS Example – Comparing Large to Small areas using 2009 5-Year

	Cook County, IL			Albany, GA			CT 303.03, Maricopa County, AZ		
	Estimate	MOE	CV	Estimate	MOE	CV	Estimate	MOE	CV
Total:	1,941,417	5,109	0.16%	36,489	853	1.42%	2,093	207	6.01%
Renter occupied:	762,028	5,506	0.44%	18,946	791	2.54%	1,218	236	11.78%
Less than \$5,000	56,326	2,039	2.20%	1,553	263	10.29%	13	20	93.52%
\$5,000 to \$9,999	65,771	1,699	1.57%	2,214	333	9.14%	92	71	46.91%
\$10,000 to \$14,999	67,399	2,311	2.08%	2,166	355	9.96%	46	41	54.18%
\$15,000 to \$19,999	61,294	2,066	2.05%	2,296	345	9.13%	134	106	48.09%
\$20,000 to \$24,999	55,439	1,619	1.78%	1,561	292	11.37%	180	102	34.45%
\$25,000 to \$34,999	98,320	2,419	1.50%	2,854	420	8.95%	383	171	27.14%
\$35,000 to \$49,999	117,466	2,333	1.21%	2,974	463	9.46%	226	130	34.97%
\$50,000 to \$74,999	122,436	2,618	1.30%	2,351	351	9.08%	0	132	x
\$75,000 to \$99,999	58,036	1,672	1.75%	557	220	24.01%	116	105	55.03%
\$100,000 to \$149,999	41,315	1,692	2.49%	256	130	30.87%	28	43	93.36%
\$150,000 or more	18,226	937	3.13%	164	99	36.70%	0	132	x

What do these examples show us?

Tracking Changes

- Do not compare ACS to Decennial Census data to track change.
- Need to study margins of error when tracking changes in 1-year estimates to distinguish actual change from sampling variability.
- Multiyear estimates have an inherent smoothing and will tend to mask rapidly developing changes.

Tracking Changes

- Change is best-measured using estimates reflecting non-overlapping time periods.
- For example, 5-year estimates for 2005 - 2009 are best compared with 5-year estimates for 2010 – 2014.
- Discontinuities between 2009 and 2010 ACS Data
 - Weighted to different decennial census data
 - Different geographies

Private Data Providers Role

- 1. Produce current year estimates and future projections.**
- 2. Adjust for margins of error, where appropriate.**
- 3. Provide their numbers and census data in easily accessible and comprehensible formats.**
- 4. Produce custom data projects**

Private Data Providers Role - Estimates

Radical methodology changes required to produce the latest estimates, due to ACS :

- From one set of base Census numbers every 10 years to 3 sets of ACS data every year.
- From single point in time (April 2000) to period averages – 3 different ones.
- Margins of error now published.
- Adding data inputs from other sources as available.

Private Data Providers - Variations

The production of demographic data is not a precise science – even though the numbers provided give the appearance of extreme precision.

There are bound to be differing results between private vendors (as well as between different data sets from the same vendor).

No way of proving which are “correct”.

Numbers provided by vendors using sound methodologies developed by expert demographers are all likely to be within a reasonable range.

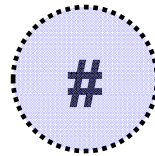
The Unobtainable Goal

The precise and current number – either a total number or a single data variable. Impossible to pinpoint in the first place and constantly changing.

#

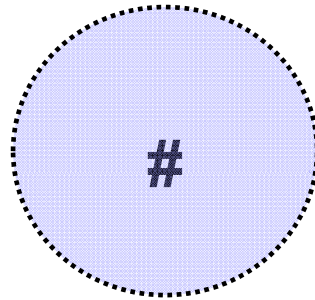
100% Count from Decennial Census

Although a 100% count, the April 2010 Census figures were not 100% correct due to those who wanted to avoid being counted, could not be located or gave incorrect answers.



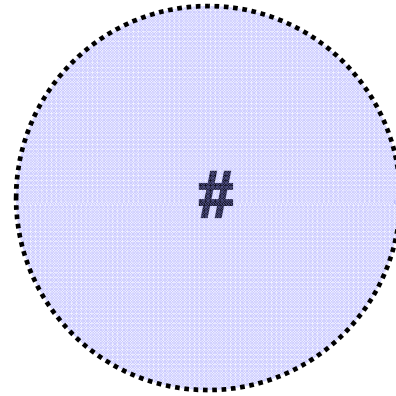
SF3 5-Year Survey Data from ACS

Less than 10% of households covered in 5 year estimates, so sampling errors occur. Numbers are 5-year averages and are not for a specific point in time.



Current year estimates based on ACS

The elapsed time between the ACS and the current year plus the need to develop single figures from multi-year averages increase the likelihood that estimates will be less precise.



Without Demographic Data

Levels of Data Accuracy

Numbers generated through sample data or update estimates are generally more accurate for:

- **Larger sample sizes (% of group surveyed)**
- **Larger area covered (geography)**
- **Wider data range (for example, \$25,000 income bands vs. \$5,000 bands)**

Accuracy of Capture Rate Results

**The fewer the layers of estimation,
the greater the likelihood of accuracy.**

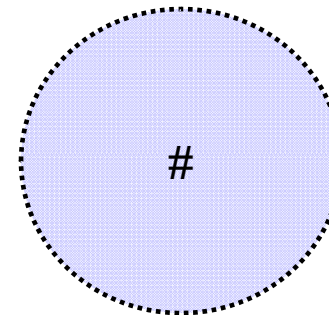
**Every step we go through to calculate a capture
rate decreases the likelihood of accuracy.**

**The overall capture rate should be given the
greatest weight.**

Accuracy of Capture Rate Results

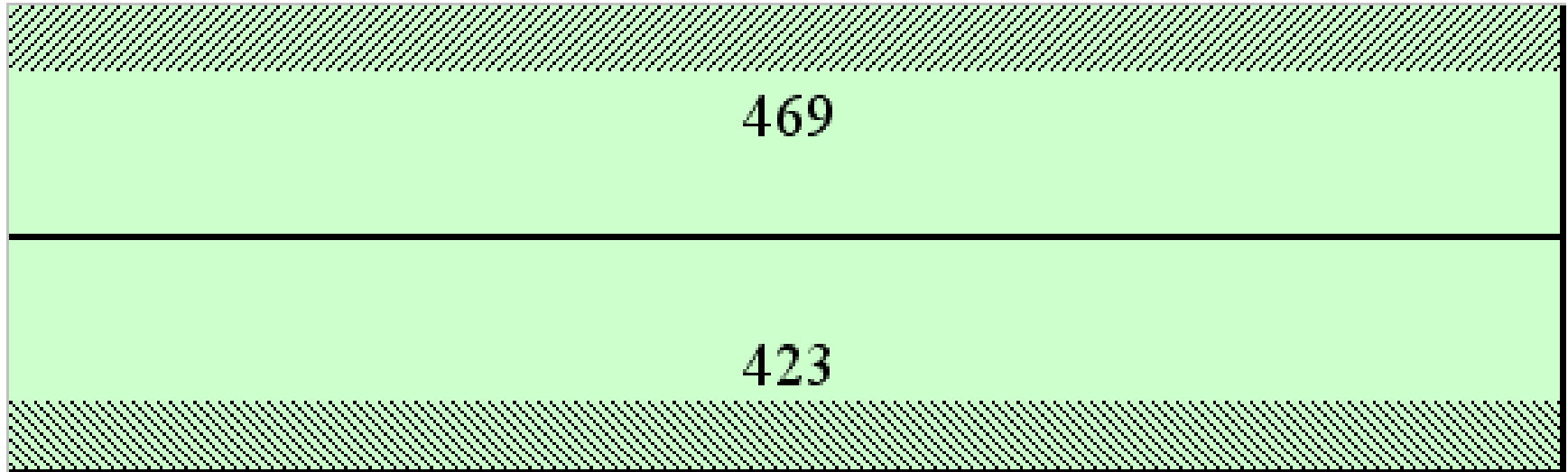
Renter Households						
Under Age 55 Years						
<i>Current Year Estimates - 2011</i>						
	1-Person Household	2-Person Household	3-Person Household	4-Person Household	5+-Person Household	Total
\$0-10,000	562	305	297	153	197	1,514
\$10,000-20,000	411	267	295	311	306	1,590
\$20,000-30,000	388	277	265	290	369	1,589
\$30,000-40,000	351	275	316	469	387	1,798
\$40,000-50,000	440	347	229	423	297	1,736
\$50,000-60,000	393	307	172	227	220	1,319
\$60,000+	<u>991</u>	<u>1,346</u>	<u>743</u>	<u>515</u>	<u>856</u>	<u>4,451</u>
Total	3,536	3,124	2,317	2,388	2,632	13,997

Numbers in green = income bands containing eligible households



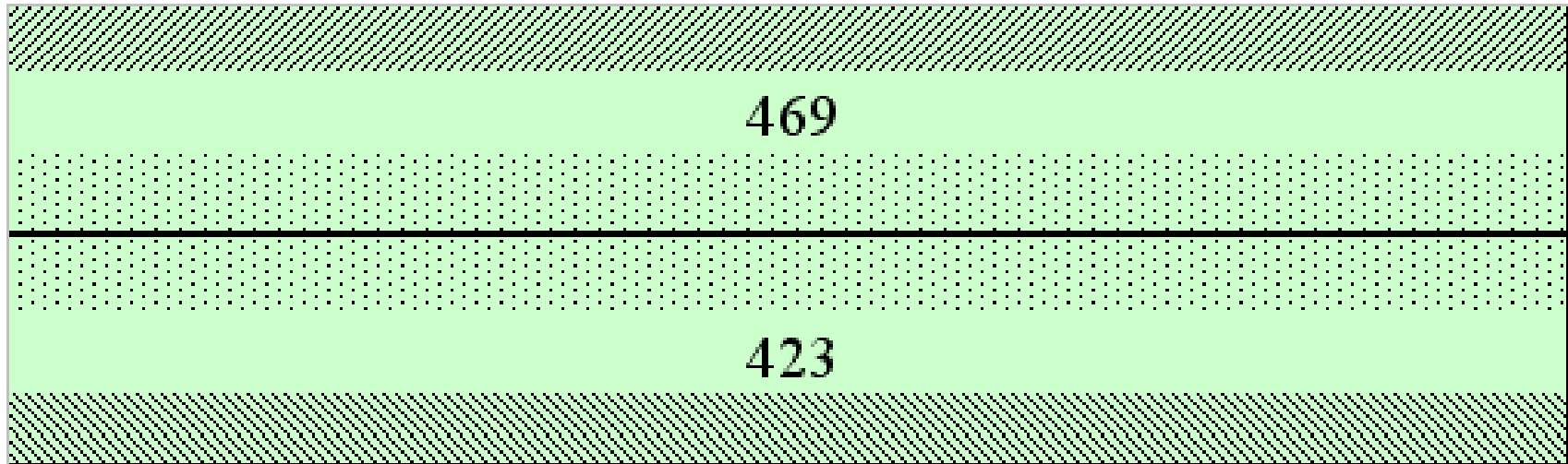
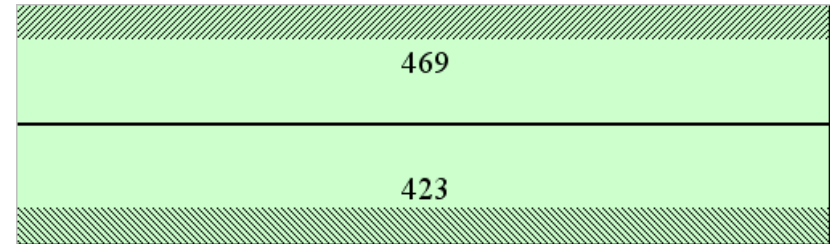
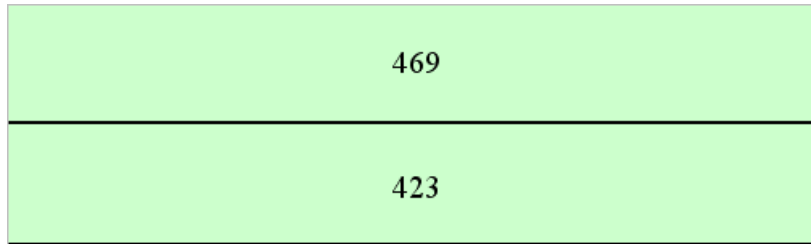
Accuracy of Capture Rate Results

469
423



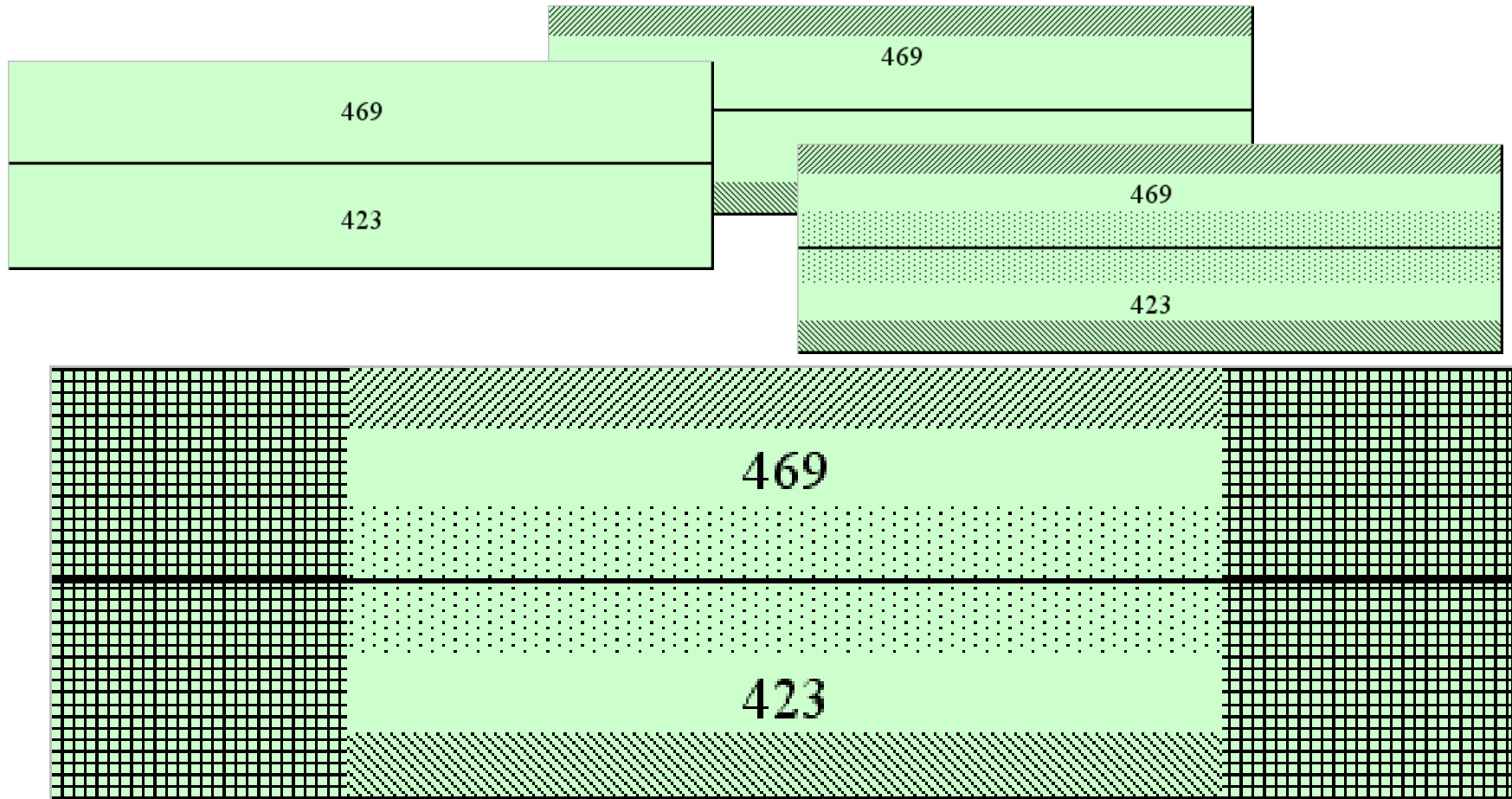
Adjusting for minimum and maximum incomes

Accuracy of Capture Rate Results



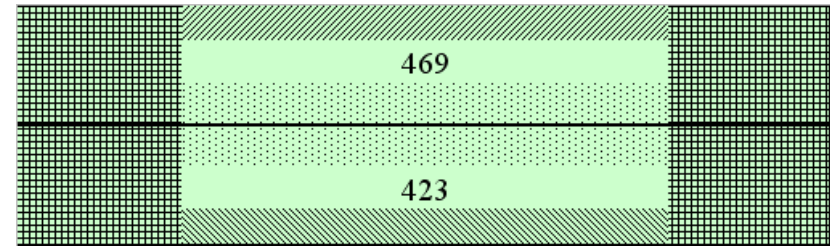
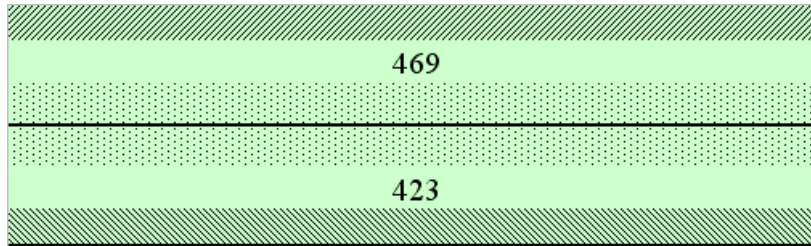
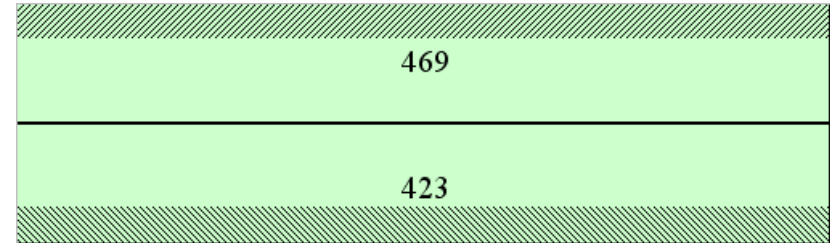
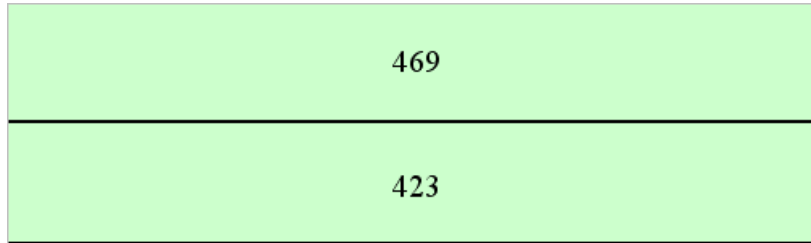
Dividing overlapping households between AMI levels

Accuracy of Capture Rate Results

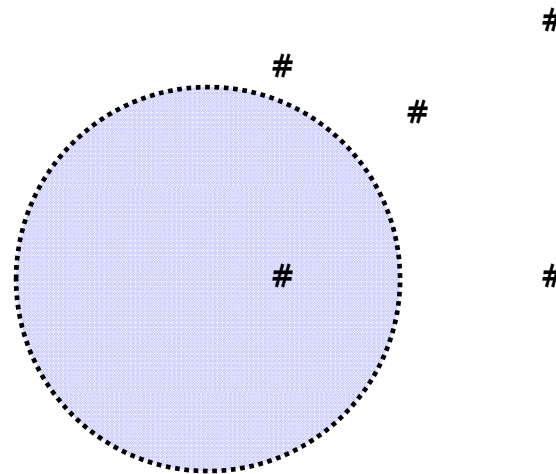


Distributing households sizes between unit type

Accuracy of Capture Rate Results



You may get lucky, but the odds aren't great.....



Capture Rate Example – Version 1

	Number Households		Number Units	Capture Rate
HHs in overall income bands	3,570		120	3.4%
HHs in adjusted income bands	2,594		120	4.6%
4-person HHs in overall income bands	892		90	10.1%
4-person HHs in adjusted income bands	624		90	14.4%
	312	← Divide between AMI levels →	312	
	Divide between unit types		Divide between unit types	
	↓ ↓ ↓		↓ ↓ ↓	
	104 104 104	← 15 units in each group →	104 104 104	Number Units
	19.2%		19.2%	Capture Rate

Capture Rate Example – Version 2

	Number Households		Number Units	Capture Rate
HHs in overall income bands	3,570		120	3.4%
HHs in adjusted income bands	2,060		120	5.8%
4-person HHs in overall income bands	892		90	10.1%
4-person HHs in adjusted income bands	479		90	18.8%
	359	← Divide between AMI levels →	120	
	Divide between unit types		Divide between unit types	
	↓ ↓ ↓		↓ ↓ ↓	
	72 216 72	← 15 units in each group →	24 72 24	Number Units
	20.9% 7.0% 20.9%		62.5% 20.8% 62.5%	Capture Rate

Capture Rate Ranges

Capture rates range from 19% in first method to between 7% and 62% in second method.

Potential for further variations due to:

- Different data source used
- Different data set requiring tenure and household size assumptions
- And, of course, market area size, which is far from an exact science as well!

Summing Up

Demographic data is not a precise science. Numbers may diverge from reality – but we can never tell exactly how much. However, they are still much better than nothing.

The larger the area/data range, the better the accuracy.

Data from different providers/datasets may show different, but valid, numbers for same area. Understanding this is key when comparing studies based on different sources.

Capture rate analysis requires applying assumptions/ratios that may move a number well away from reality.

Differing methodologies among market analysts may show different results for the same area and project – even with same data.

Figuring Out Appropriate Uses

Demographic and capture rate analysis should be regarded as just one piece of a large puzzle.

Now we understand the limitations more clearly, are we obligated to reconsider how we use the data?

We can calculate very detailed capture rate estimates
– but should we?

Would a range of capture rates for each property be fairer for the comparison of competing projects?

What level of variation should be considered “margins of error” in capture rates?

Is there a case for doing away with scoring capture rates and setting thresholds?